

# 1. Pendahuluan

## 1.1 Latar belakang masalah

*Rating* film merupakan suatu penilaian suka atau tidaknya *user* terhadap film tertentu (*range* penilaian adalah *integer* 1-5). Dengan adanya penilaian *rating* tersebut berguna bagi sebuah perusahaan *rental* film *online* untuk meningkatkan pemasaran yaitu dengan menawarkan film yang diperkirakan disukai oleh pelanggan dengan melihat *rating* yang telah diberikan *user* pada film. Pada kenyataannya tidak semua film di beri *rating* oleh *user* karena terlalu banyaknya film yang tersedia, sehingga perusahaan akan kesulitan untuk menentukan film mana yang harus ditawarkan ke pelanggan.

Untuk mengatasi hal di atas perlu di buat sebuah sistem yang dapat merekomendasikan kepada pelanggan, film-film yang diprediksi disukai oleh pelanggan [2] berdasarkan beberapa penilaian *rating* oleh pelanggan yang telah ada. Sistem rekomendasi tersebut adalah suatu sistem yang dapat mempermudah bagi pelanggan untuk memperoleh film yang mereka sukai tanpa harus menyeleksi atau memilah semua judul, mengingat jumlah film yang tersedia sangat banyak sehingga tidak memungkinkan bagi tiap *user* untuk memeriksa semua film yang ada [8].

Pada tugas akhir ini digunakan metode *Counterpropagation Network* untuk membuat sistem rekomendasi yang dapat memprediksi nilai *rating* film. Input dan target sistem adalah berupa nilai *rating* yang diberikan user terhadap suatu film tertentu. Data yang digunakan disediakan pada website [www.netflixprize.com](http://www.netflixprize.com).

*Counterpropagation Network* pertama kali diperkenalkan oleh Robert Hecht-Nielsen pada tahun 1986. Proses pembelajaran pada *CPN* terdiri dari 2 *fase*:

1. Layer Kohonen Self Organized Maps (unsupervised)
2. Layer Grossberg (Supervised)

Alasan menggunakan metode ini dikarenakan setelah data yang ada diamati lebih lanjut, penulis berpendapat bahwa untuk menyelesaikan masalah tersebut diperlukan suatu metode *clustering* untuk mengurangi dimensi data yang sangat besar. Oleh karena itu digunakanlah metode *Kohonen* yang merupakan *competitive layer* sebagai peng-*cluster* data *input*, sedangkan untuk hasil prediksi itu sendiri digunakan metode *Grosberg* yang akan mengoreksi kesalahan pada output sistem terhadap target yang sebenarnya sehingga akan mengurangi tingkat kesalahan atau *error*.

## 1.2 Perumusan masalah

Dari gambaran di atas maka dapat diambil perumusan masalah pokok sebagai berikut:

1. Bagaimana menerapkan metode *Counterpropagation Network* untuk dapat memprediksi berapa skala *rating* yang akan diberikan *user* pada film tertentu.
2. Bagaimana akurasi metode *Counterpropagation Network* dalam menganalisa kasus prediksi *rating* film.

## 1.3 Tujuan

Tujuan dari tugas akhir ini adalah :

1. Dapat memprediksi berapa skala *rating* yang akan diberikan *user* pada film tertentu dengan menggunakan metode *Counterpropagation Network*.
2. Menganalisa *performansi* metode *Counterpropagation Network* dalam kasus prediksi *rating* film dengan skala *RMSE (Root Mean Square Error)*.

## 1.4 Batasan Masalah

Batasan masalah dalam ”*Prediksi Rating Film Menggunakan Metode Counterpropagation Network*” diantaranya :

1. Sebelum data masuk ke sistem, data telah di olah dengan menggunakan *matrix factorization* sehingga data menjadi *matrix input* dan target *JST*.
2. Data yang digunakan merupakan data dari situs [www.netflixprize.com](http://www.netflixprize.com) yang dikumpulkan mulai Oktober 1998 sampai Desember 2005.
3. Dari data yang tersedia pada file *mv\_<movie id>.txt* hanya 1.18 % dari keseluruhan data yang ada dan yang hilang mencapai 98.82 %, sehingga diambil 20-80 % dari data yang tersedia sebagai data *training*. Tujuan pengambilan data *training* yang bervariasi tersebut adalah untuk mengetahui apakah dengan peningkatan jumlah data *training* akan memperbaiki tingkat kesalahan pada saat *testing* maupun *predicting*, mengingat kondisi data yang tersedia terlalu sedikit. Untuk data *validating* diambil 10 % dari file *probe.txt*, sedangkan untuk data *testing* diambil dari file *probe.txt* selain data *validating*. Sedangkan untuk data yang akan diprediksi nilai *rating*nya diambil dari file *qualifying.txt*.
4. Prediksi yang dilakukan tidak melibatkan secara langsung faktor-faktor penentu kesukaan pelanggan terhadap film seperti jenis film (*horror*, *aksi*, *serial dll*), film yang sedang *ngetren*, aktor ataupun aktris pada film, produser, tanggal pembuatan film, kapan film dirating oleh *user* dll. Data yang digunakan dianggap sudah mengandung faktor-faktor tersebut secara implisit.
5. *RMSE (root mean square error)* data prediksi dilakukan oleh pihak [www.netflixprize.com](http://www.netflixprize.com) sebagai penyedia data karena *target rating* tidak disediakan. Caranya yaitu dengan meng-*upload* hasil prediksi sistem ke pihak [www.netflixprize.com](http://www.netflixprize.com), *RMSE* hasil *predicting* akan dikirim melalui *email* yang telah didaftarkan.

## 1.5 Metodologi penyelesaian masalah

Metodologi penelitian penyusunan tugas akhir ini adalah sebagai berikut:

1. **Studi Literatur**, mempelajari buku, jurnal, dan referensi lain mengenai *neural network Kohonen*, *Self Organising Maps*, dan *Counterpropagation Network*.
2. **Pendalaman Materi**, mempelajari dan memahami konsep dan teori yang mendukung proses pembuatan, pengembangan dan penyempurnaan suatu sistem prediksi yang baik.
3. **Pengumpulan Data**, meliputi pengumpulan dataset berupa data nilai rating film oleh user/pelanggan perusahaan netflix yang dikumpulkan mulai Oktober 1998 sampai Desember 2005.
4. **Analisa dan Perancangan Metode**, melakukan analisis dan perancangan implementasi perangkat lunak untuk memprediksi nilai skala *rating* film oleh user/ pelanggan menggunakan Matlab.
5. **Implementasi dan Pengujian**, mengimplementasikan perangkat lunak yang akan dibangun dan melakukan pengujian implementasi dengan data yang ada.
6. **Pengambilan Kesimpulan**, Kesimpulan diambil berdasarkan pada analisis hasil pengujian.